Applicant hereby submits Exhibit A, which illustrates the homology of the claimed marker UT116 and PSCA. As shown by Exhibit A, UT116 is identical to PSCA. Further, Applicant submits Exhibit B, which is an article, (Amara, N., *et al.*, "Prostate Stem Cell Antigen Is Overexpressed in Human Transitional Cell Carcinoma", *Cancer Research,* **61**:4660-4665 (2001)), showing the link between PSCA (UT116) and bladder cancer. Specifically, it was shown that PSCA (UT116) was over-expressed in virtually all non-muscle invasive bladder tumors and that these results support PSCA (UT116) as a potential diagnostic target in bladder cancer (p. 4660, line 48 through p. 4661, line 7). Therefore, it is respectfully requested that this rejection be withdrawn.

The Examiner also rejects these claims stating that while enabling for the polynucleotides of SEQ ID NOS. 1-12, there is not enablement for polynucleotides having 95% identity to these sequences.

Therefore, Applicant submits the software manual to the Wisconsin Sequence Analysis program, Version 8, publicly available from Genetics Computer Group, Madison, WI, as Exhibit C. Support for this submission is found on page 12, beginning on line 10. The manual provides the algorithm, parameters, parameter values and other information necessary to, accurately and consistently, calculate the percent identity. This manual indicates on pages 5-21, *inter alia,* that the software used the local homology algorithm of Smith and Waterman (Advances in Applied Mathematics 2; 482-489 (1981)).

CONCLUSION

In view of the aforementioned remarks, Applicant respectfully submits that the above-referenced application is now in a condition for allowance and Applicant

respectfully requests that the Examiner withdraw all outstanding objections and rejections and passes the application to allowance.

23492

ABBOTT LABORATORIES
Telephone: (847) 935-7550
Facsimile: (847) 938-2623

Respectfully submitted,
P. A. Billing-Medel, et al.

Mimi C. Goller
Registration No. 39,046
Attorney for Applicants

EXHIBIT A

Alignments

>gi|9367211|emb|AJ297436.1|HSA297436 Homo sapiens mRNA for prostate stem cell antigen (PSCA gene)
Length = 946

Score = 1816 bits (916), Expect = 0.0
Identities = 919/920 (99%)
Strand = Plus / Plus

```
Query: 41  acagcccaccagtgaccatgaaggctgtgctgcttgccctgttgatggcaggcttggccc 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 3   acagcccaccagtgaccatgaaggctgtgctgcttgccctgttgatggcaggcttggccc 62

Query: 101 tgcagccaggcactgccctgctgtgctactcctgcaaagcccaggtgagcaacgaggact 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 63  tgcagccaggcactgccctgctgtgctactcctgcaaagcccaggtgagcaacgaggact 122

Query: 161 gcctgcaggtggagaactgcacccagctgggggagcagtgctggaccgcgcgcatccgcg 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 123 gcctgcaggtggagaactgcacccagctgggggagcagtgctggaccgcgcgcatccgcg 182

Query: 221 cagttggcctcctgaccgtcatcagcaaaggctgcagcttgaactgcgtggatgactcac 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 183 cagttggcctcctgaccgtcatcagcaaaggctgcagcttgaactgcgtggatgactcac 242

Query: 281 aggactactacgtgggcaagaagaacatcacgtgctgtgacaccgacttgtgcaacgcca 340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 243 aggactactacgtgggcaagaagaacatcacgtgctgtgacaccgacttgtgcaacgcca 302

Query: 341 gcggggcccatgccctgcagccggctgccgccatccttgcgctgctccctgcactcggcc 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 303 gcggggcccatgccctgcagccggctgccgccatccttgcgctgctccctgcactcggcc 362
```

```
Query: 401 tgctgctctggggacccggccagctataggctctgggggggccccgctgcagcccacactg 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 363 tgctgctctggggacccggccagctataggctctgggggggccccgctgcagcccacactg 422


Query: 461 ggtgtggtgccccaggcctctgtgccactcctcacagacctggcccagtgggagcctgtc 520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 423 ggtgtggtgccccaggcctctgtgccactcctcacagacctggcccagtgggagcctgtc 482


Query: 521 ctggttcctgaggcacatcctaacgcaagtctgaccatgtatgtctgcacccctgtcccc 580
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct: 483 ctggttcctgaggcacatcctaacgcaagtctgaccatgtatgtttgcacccctgtcccc 542


Query: 581 caccctgaccctcccatggccctctccaggactcccacccggcagatcagctctagtgac 640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 543 caccctgaccctcccatggccctctccaggactcccacccggcagatcagctctagtgac 602


Query: 641 acagatccgcctgcagatggcccctccaaccctctctgctgctgtttccatggcccagca 700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 603 acagatccgcctgcagatggcccctccaaccctctctgctgctgtttccatggcccagca 662


Query: 701 ttctccacccttaaccctgtgctcaggcacctcttcccccaggaagccttccctgcccac 760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 663 ttctccacccttaaccctgtgctcaggcacctcttcccccaggaagccttccctgcccac 722


Query: 761 cccatctatgacttgagccaggtctggtccgtggtgtcccccgcacccagcaggggacag 820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 723 cccatctatgacttgagccaggtctggtccgtggtgtcccccgcacccagcaggggacag 782


Query: 821 gcactcaggagggcccagtaaaggctgagatgaagtggactgagtagaactggaggacaa 880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 783 gcactcaggagggcccagtaaaggctgagatgaagtggactgagtagaactggaggacaa 842


Query: 881 gagtcgacgtgagttcctgggagtctccagagatggggcctggaggcctggaggaagggg 940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 843 gagtcgacgtgagttcctgggagtctccagagatggggcctggaggcctggaggaagggg 902


Query: 941 ccaggcctcacattcgtggg 960
           ||||||||||||||||||||
Sbjct: 903 ccaggcctcacattcgtggg 922
```

[CANCER RESEARCH 61, 4660–4665, June 15, 2001]

Exhibit B

Advances in Brief

Prostate Stem Cell Antigen Is Overexpressed in Human Transitional Cell Carcinoma[1]

Nordine Amara,[2] Ganesh S. Palapattu,[2] Matthew Schrage, Zhennan Gu, George V. Thomas, Fred Dorey, Jonathan Said, and Robert E. Reiter[3]

Departments of Urology [N. A., G. S. P., Z. G., R. E. R.], Pathology [M. S., G. V. T., J. S.], and Orthopedics [F. D.], and Jonsson Cancer Center [N. A., G. S. P., Z. G., R. E. R.], University of California-Los Angeles School of Medicine, Los Angeles, California 90095

Abstract

Prostate stem cell antigen (PSCA), a homologue of the Ly-6/Thy-1 family of cell surface antigens, is expressed by a majority of human prostate cancers and is a promising target for prostate cancer immunotherapy. In addition to its expression in normal and malignant prostate, we recently reported that PSCA is expressed at low levels in the transitional epithelium of normal bladder. In the present study, we compared the expression of PSCA in normal and malignant urothelial tissues to assess its potential as an immunotherapeutic target in transitional cell carcinoma (TCC). Immunohistochemical analysis of PSCA protein expression was performed on tissue sections from 32 normal bladder specimens, as well as 11 cases of low-grade transitional cell dysplasia, 21 cases of carcinoma *in situ* (CIS), 38 superficial transitional cell tumors (STCC, stages T_a-T_1), 65 muscle-invasive TCCs (ITCCs, stages T_2-T_4), and 7 bladder cancer metastases. The level of PSCA protein expression was scored semiquantitatively by assessing both the intensity and frequency (*i.e.*, percentage of positive tumor cells) of staining. We also examined PSCA mRNA expression in a representative sample of normal and malignant human transitional cell tissues. In normal bladder, PSCA immunostaining was weak and confined almost exclusively to the superficial umbrella cell layer. Staining in CIS and STCC was more intense and uniform than that seen in normal bladder epithelium ($P < 0.001$), with staining detected in 21 (100%) of 21 cases of CIS and 37 (97%) of 38 superficial tumors. PSCA protein was also detected in 42 (65%) of 65 of muscle-invasive and 4 (57%) of 7 metastatic cancers, with the highest levels of PSCA expression (*i.e.*, moderate-strong staining in >50% of tumor cells) seen in 32% of invasive and 43% of metastatic samples. Higher levels of PSCA expression correlated with increasing tumor grade for both STCCs and ITCCs ($P < 0.001$). Northern blot analysis confirmed the immunohistochemical data, showing a dramatic increase in PSCA mRNA expression in two of five muscle-invasive transitional cell tumors when compared with normal samples. Confocal microscopy demonstrated that PSCA expression in TCC is confined to the cell surface. These data demonstrate that PSCA is overexpressed in a majority of human TCCs, particularly CIS and superficial tumors, and may be a useful target for bladder cancer diagnosis and therapy.

Introduction

TCC[4] of the bladder poses a significant worldwide clinical problem, with an estimated 54,200 new cases and 12,100 associated deaths reported in the United States in 1999 alone (1). The majority of patients present initially with superficial disease limited to the urothelium or lamina propria of the bladder wall. Such lesions are often amenable to transurethral resection. Intravesical immunotherapy with BCG may prevent recurrence or progression of high-risk tumors (*e.g.*, high grade or T_1) and CIS (2). However, despite these efforts, ~50% of superficial tumors will continue to recur and as many as 30% will progress to muscle-invasive disease (3). Although radical cystectomy can salvage many patients with muscle-invasive cancers, a significant number go on to die from metastatic disease, for which there is currently no effective treatment. These data underscore the urgent necessity for better diagnostic and treatment strategies for superficial and invasive bladder cancers.

Given the sensitivity of bladder cancer to BCG immunotherapy, there is a particular need to identify bladder cancer antigens for cellular and monoclonal antibody-based targeted immunotherapies. EGFR, for example, is overexpressed by a significant percentage of muscle-invasive bladder cancers (4). A recent study demonstrated that monoclonal antibody directed against EGFR slowed growth of a human transitional cell cancer in an orthotopic mouse model (5). Similarly, new bladder cancer markers have been identified that not only demonstrate high specificity for TCC but that also show early promise as a clinical tool. One such marker, uroplakin II, is an urothelium-specific differentiation antigen that is expressed by ~40% of TCCs (6). Detection of uroplakin-positive cells in human sera has been associated with metastatic spread of bladder cancer cells and may identify patients with micrometastatic spread prior to undergoing cystectomy (7).

PSCA is a glycosylphosphatidylinositol (GPI)-anchored 123-amino-acid glycoprotein related to the Ly-6/Thy-1 family of cell surface antigens (8). PSCA expression in normal tissues is largely prostate-specific, but we recently reported finding PSCA transcripts and protein in transitional epithelium of the bladder and neuroendocrine cells of the stomach (9). *In situ* hybridization and IHC analyses demonstrated PSCA expression in more than 80% of local and 100% of bone-metastatic prostate cancer specimens (9). Importantly, the intensity of PSCA expression increased with tumor grade and stage, which suggests its potential as an immunotherapeutic target for high-risk and metastatic prostate cancer. Supporting this hypothesis, we recently demonstrated that monoclonal antibodies against PSCA can inhibit tumor growth and metastasis formation and can prolong survival in mice bearing human prostate cancer xenografts (10). Also, Dannull *et al.* (11) recently reported that a PSCA-derived peptide could elicit a PSCA-specific T-cell response in a patient with metastatic prostate cancer.

Because PSCA is present at low levels in normal bladder, we asked whether PSCA is expressed in TCC. We also determined whether PSCA is overexpressed in bladder cancer compared with normal bladder and whether the level of expression correlates with bladder cancer stage or grade. We demonstrate that PSCA is expressed by a majority of both muscle-invasive and superficial tumors. Moreover,

Received 2/2/01; accepted 4/27/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported in part by grants from the NIH K08 CA74169, the Cancer Research Institute, and CaPCURE.

[2] N. A. and G. S. P. contributed equally to this work.

[3] To whom requests for reprints should be addressed, at the Department of Urology, UCLA, 66-134 Center for the Health Sciences, 10833 Le Conte Avenue, Los Angeles, CA 90095-1738. E-mail: rreiter@mednet.ucla.edu.

[4] The abbreviations used are: TCC, transitional cell carcinoma; PSCA, prostate stem cell antigen; ITCC, invasive TCC; STCC, superficial TCC; CIS, carcinoma *in situ*; IHC, immunohistochemical; BCG, Calmette-Guerin bacillus; EGFR, epidermal growth factor receptor.

PSCA is overexpressed in virtually all nonmuscle invasive bladder tumors and in >30% of invasive and metastatic cancers. As with prostate cancer, expression increases with increasing tumor grade. Interestingly, the overexpression of PSCA that we observed in TCC is quantitatively more than that seen in prostate cancer with respect to tumor stage. These results support PSCA as a potential diagnostic and/or therapeutic target in bladder cancer.

Materials and Methods

Tissue Samples

All of the tissue specimens were obtained with permission from the Human Tissue Resources Committee of the Department of Pathology at University of California-Los Angeles Medical Center. One hundred seventy-four formalin-fixed, paraffin-embedded human bladder tissues were obtained from 135 different patients. Blocks were then cut into 4-μm sections and mounted on charged slides in the usual fashion. H&E-stained sections of the neoplasms were graded by an experienced urological pathologist according to the criteria set forth in the International Histological Classification of Tumors. Staging was performed based on the 1997 Tumor-Node-Metastasis classification system. The tissue samples consisted of 32 normal bladder samples, 11 cases of low-grade transitional cell dysplasia, 21 cases of CIS, 38 cases of STCC (stages T_a–T_1,), 65 cases of muscle-ITCC, (stages T_2–T_4), and 7 metastases (6 lymph nodes, 1 lung). Normal bladder samples were obtained from cystectomy specimens of patients with focal muscle-invasive disease and no preoperative cystoscopic or cytological evidence of CIS. Areas histologically free of tumor were taken as normal bladder tissue. Of the specimens in the STCC group, 13 were grade II and 25 were grade III. One sample in the ITCC group was grade I, 12 were grade II, and 53 were grade III.

Immunohistochemistry

Methodology. Specimens were stained using modifications of an immunoperoxidase technique previously described (12). Briefly, antigen retrieval was performed on paraffin sections using a commercial steamer and 0.01 M citrate buffer (pH 6.0). Slides were then washed in PBS and incubated with normal horse serum, diluted 1:20, for 10 min. PSCA monoclonal antibody 1G8 was generated in the CellPharm System 100 as described previously (9). Monoclonal antibodies to PSCA were diluted 1:20 in PBS. After 50 min of incubation with the primary antibody, slides were treated sequentially with rabbit antimouse immunoglobulin, swine antirabbit immunoglobulin, and rabbit antiswine immunoglobulin (all biotin conjugated). Slides were then incubated with streptavidin-peroxidase, and antibody localization was performed using the diaminobenzidene reaction. Positive and negative controls were performed on tissues obtained from mouse xenograft tumors that were derived from the human prostate cancer cell lines LAPC-9 and PC3, respectively. Negative controls for each stained section consisted of substitution of the primary antibody by a non-cross-reacting isotype-matched monoclonal antibody.

Scoring Methods. Histopathological slides of the clinical specimens were read and scored by two pathologists (G. V. T., and J. S.) in a blinded fashion. There was a >90% inter- and intraobserver agreement. IHC intensity was graded on a scale of 0 to 3+ (0, no staining; 1+, mildly intense; 2+, moderately intense; 3+, severely intense). Staining density was quantified as the percentage of cells staining positive with the primary antibody, as follows: 0 = no staining, 1 = superficial staining, 2 = positive staining in <25% of the sample; 3 = positive staining in 25–50% of the sample; 4 = positive staining in >50% of the specimen; and 5 = positive staining throughout the sample. Intensity score (0 to 3+) was multiplied by the density score (0–5) to calculate an overall score (0–15). In this way, we were able to differentiate specimens that may have had focal areas of increased staining from those that had diffuse areas of increased staining (13). The overall score for each specimen was then categorically assigned to one of the following groups: 0–3 (little to no immunoreactivity), 4–7 (moderate immunoreactivity), and 8–15 (strong immunoreactivity). The above groupings attempt to stratify the overall PSCA immunoreactivity of each specimen for the purpose of comparison. Overexpression was considered to be present when the overall score was ≥8, indicating that 50–100% of cells had moderate or intense staining. Within any given slide, adjacent areas of normal transitional epithelium were used as positive internal controls.

Northern Blot Analysis. Fresh human tissues were obtained at the time of surgery, and RNA was prepared according to the manufacturer's recommendations (Biotecx, Houston, TX). Nine samples were obtained: four benign and five malignant transitional cell tissues. All five malignant tissues were from patients with muscle-invasive disease. Normal bladder RNA was obtained from benign areas of cystectomy specimens of patients with focal muscle-ITCC. In none of these patients was there evidence, either pre- or postoperatively, of CIS. Northern blot analysis was then performed as described previously (14). Briefly, 10 μm of RNA from each sample were loaded onto a 1% agarose gel. Equal loading and RNA integrity was confirmed by ethidium bromide staining. After electrophoresis, RNA samples were transferred onto a nitrocellulose membrane. The PSCA probe was prepared from a cDNA fragment using random oligonucleotide primers (Amersham) and was labeled with [^{32}P]dCTP. Equal loading of RNA was also confirmed by hybridization with an actin probe. The human prostate cancer cell lines LAPC-4 and LNCaP were used as positive and negative controls, respectively. Preparation of total RNA from cell lines was performed according to the manufacturer's instructions (Biotecx). The human bladder cancer cell line HT1376 was also included in the analysis.

Confocal Microscopy. Human bladder cancer cell line HT1376 cells were washed with PBS containing 1 mM $MgCl_2$ and 0.1 mM $CaCl_2$ (PBS/CM). Cells were incubated with the primary antibody, murine monoclonal anti-PSCA antibody 1G8, and subsequently with a secondary antibody, fluorescein-conjugated FITC goat antimouse IgG (1:200; Jackson Immunoresearch, West Grove, PA). The cells were then fixed with 2% paraformaldehyde in PBS/CM for 30 min. Next, the cells were incubated with 50 mM NH_4Cl in PBS/CM for 10 min at room temperature and permeabilized for 10 min with 0.075% (w/v) saponin in PBS/CM containing 0.2% BSA. Nuclei were counterstained with propidium iodide (1 μg/ml; Sigma Chemical Co.). Optical sections were obtained by laser confocal microscopy.

Statistical Analysis. Overall scores of PSCA staining were evaluated with respect to tissue histology using a one-way ANOVA-Sidak multiple comparison procedure, using overall score as the dependent variable and the histopathological subgroup as the factor. Table 3 was constructed in this manner. To evaluate the relationship between tumor grade and overall score within the subgroups STCC and ITCC, we performed a two-way ANOVA using the overall score as the dependent variable and the grade and subgroup as the two factors. The results of this analysis can be seen in Table 2. For all analyses, $P < 0.05$ was considered statistically significant.

Results

PSCA Protein Expression in Normal, Dysplastic, and Malignant Urothelium. The expression pattern of PSCA in normal bladder tissues and TCCs of different stages and grades is summarized in Tables 1 and 2 and illustrated in Fig. 1, *A–H*. Thirty of 32 normal bladder specimens stained positively for PSCA. In these samples, PSCA expression was generally weak (*i.e.*, 1+) and was localized almost exclusively to the superficial umbrella cell layer (*i.e.*, density score of 1; Fig. 1*A*). The mean intensity, density, and overall staining scores from normal tissues were 1.0, 1.0, and 1.1, respectively (Table 1). Expression of PSCA in normal bladders was significantly less than that found in CIS and both invasive and superficial tumors ($P = 0.001$; Table 3).

PSCA expression was detected in all of the 11 low-grade dysplasia specimens surveyed (Table 1; IHC data not shown). The mean intensity, density, and overall score for these samples was 1.8, 2.0, and 3.6, respectively. Although PSCA staining intensity and density tended to be higher in dysplastic than in normal bladder, this was not statistically significant ($P = 0.782$; Table 3). In contrast, PSCA expression in CIS was intense and homogeneous in all but two specimens (19 of 21), with a mean intensity, density, and overall score of 2.7, 4.9, and 13.2, respectively (Table 1). Whereas PSCA staining in normal and dysplastic bladders was confined to the most superficial layers of urothelium, staining in CIS was detected in all of the neoplastic

Table 1 *Comparison of IHC staining intensity, density, and overall score by stage*

	Intensity					Density							Overall Score			
	0	1+	2+	3+	Mean ± SD	0	1	2	3	4	5	Mean ± SD	0–3	4–7	8–15	Mean ± SD
Normal, n = 32 (100%)	2 (6)	28 (88)	2 (6)	0 (0)	1.0 + 0.4	2 (6)	28 (88)	2 (6)	0 (0)	0 (0)	0 (0)	1.0 + 0.4	32 (100)	0 (0)	0 (0)	1.1 + 0.8
Dysplasia, n = 11 (100%)	0 (0)	5 (45)	3 (27)	3 (27)	1.8 + 0.9	0 (0)	0 (0)	11 (100)	0 (0)	0 (0)	0 (0)	2.0 + 0.0	5 (45)	6 (55)	0 (0)	3.6 + 1.7
CIS, n = 21 (100%)	0 (0)	2 (10)	3 (14)	16 (76)	2.7 + 0.7	0 (0)	0 (0)	1 (5)	0 (0)	0 (0)	20 (95)	4.9 + 0.7	1 (5)	1 (5)	19 (90)	13.2 + 3.7
Superficial TCC																
T_a, n = 19 (100%)	0 (0)	4 (21)	11 (58)	4 (21)	2.0 + 0.7	0 (0)	0 (0)	3 (16)	2 (11)	5 (26)	9 (47)	4.1 + 1.1	4 (21)	1 (5)	14 (74)	8.6 + 4.2
T_1, n = 19 (100%)	1 (5)	2 (11)	8 (42)	8 (42)	2.2 + 0.9	1 (5)	0 (0)	0 (0)	1 (5)	1 (5)	16 (84)	4.6 + 1.2	1 (5)	3 (16)	15 (79)	10.7 + 4.5
Invasive TCC																
T_2, n = 19 (100%)	6 (32)	4 (21)	5 (26)	4 (21)	1.4 + 1.2	6 (32)	0 (0)	6 (32)	4 (21)	2 (11)	1 (5)	1.9 + 1.6	9 (47)	6 (32)	4 (21)	4.1 + 4.1
T_3, n = 39 (100%)	15 (39)	6 (15)	6 (15)	12 (31)	1.4 + 1.3	15 (39)	0 (0)	3 (8)	6 (15)	6 (15)	9 (23)	2.4 + 2.1	17 (44)	9 (23)	13 (33)	5.5 + 5.6
T_4, n = 7 (100%)	2 (29)	0 (0)	1 (14)	4 (57)	2.0 + 1.4	2 (29)	0 (0)	1 (14)	3 (43)	0 (0)	1 (14)	2.3 + 1.8	2 (29)	1 (14)	4 (57)	6.6 + 5.5
Metastasis, n = 7 (100%)	3 (43)	1 (14)	0 (0)	3 (43)	1.4 + 1.5	3 (43)	0 (0)	1 (14)	0 (0)	1 (14)	2 (29)	2.3 + 2.4	4 (57)	0 (0)	3 (43)	6.3 + 7.3

urothelial cell layers (Fig. 1*B*). Overall, CIS expressed significantly higher levels of PSCA than any other specimen category in this study ($P < 0.05$, compared with normal bladder, dysplasia, STCC, ITCC, and metastases; Table 3).

Nonmuscle invasive tumors (stages T_a and T_1) also displayed an overall strong positivity for PSCA expression (Fig. 1*C*). Twenty-nine (76%) of 38 samples overexpressed PSCA, defined as an overall score of ≥8 (Table 1). The mean intensity, density, and overall scores for T_a tumors were 2.0, 4.1, and 8.6, respectively, whereas T_1 lesions scored 2.2, 4.6, and 10.7, respectively. As a group, superficial tumors stained more strongly than normal, dysplastic, and invasive samples ($P = 0.001$; Table 3). PSCA expression increased significantly with worsening tumor grade in nonmuscle invasive tumors (Table 2). Grade 2 lesions had a mean overall staining score of 6.8 compared with a mean score of 11.2 for Grade 3 tumors ($P < 0.001$).

We detected PSCA expression in 42 (65%) of 65 muscle-invasive tumors, of which 21 (32%) exhibited the highest levels of expression (*i.e.*, overall score, ≥8). Locally advanced tumors tended to express higher levels of PSCA than organ-confined lesions, although this trend did not reach statistical significance. T_2 tumors had a mean overall score of 4.1, whereas T_3 and T_4 tumors had mean intensity *versus* density scores of 5.5 and 6.6, respectively (Table 1; Fig. 1, *D–F*). As with superficial tumors, PSCA expression in muscle-invasive cancers increased significantly with tumor grade (Table 2). Eight (67%) of 12 grade 2 tumors displayed no detectable PSCA expression, whereas only 14 (27%) of 53 grade 3 lesions showed no expression. The mean overall score for grade 2 muscle-invasive tumors was 2.1, compared with a score of 5.9 for grade 3 tumors ($P < 0.001$). Several poorly differentiated muscle-invasive TCCs contained foci of squamous cell differentiation. PSCA expression in these cases was particularly prominent, as depicted in Fig. 1*F*. We also examined seven bladder cancer metastases. Four (57%) stained positive for PSCA (Fig. 1*G*), three of which were in the strongest immunoreactivity category (Table 1). The mean intensity, density, and overall scores for metastases were 1.4, 2.3, and 6.3, respectively. These results demonstrate that PSCA is expressed by a majority of TCCs. Expression is greatest in CIS and superficial tumors but is also extremely high in >30% of muscle-invasive and metastatic lesions. Importantly, within the subgroups STCC and ITCC, higher levels of PSCA expression correlate significantly with increasing tumor grade.

PSCA Protein Is Expressed on the Cell Surface of Bladder Cancer Cells. To confirm that PSCA localizes to the cell surface of bladder cancer cells, we stained nonpermeabilized HT1376 bladder cancer cells and examined them by confocal microscopy. PSCA monoclonal antibody 1G8 detected PSCA expression in a subpopulation of these cells. As seen in Fig. 1*H*, PSCA protein expression is localized to the cell surface.

PSCA mRNA Expression in Normal and Malignant Urothelium. Northern blot analysis was performed on four normal and five invasive transitional cell cancer samples (Fig. 1*I*). As predicted, low levels of PSCA mRNA expression were detected in three of four normal bladder specimens. PSCA mRNA was also detected in three (60%) of five muscle-invasive tumors and an invasive transitional cell cancer cell line (HT1376). Consistent with the IHC findings, two of these invasive cancers (*i.e.*, 40%) overexpressed PSCA when compared with normal bladder. The level of PSCA expression in these two cases was similar to that seen in LAPC 9, a prostate cancer xenograft. The corresponding IHC stain for a patient with a muscle-invasive transitional cell tumor of the renal pelvis is shown in Fig. 1*E* (*lane with asterisk*) and confirms the intense expression detected by Northern analysis. Of all of the tissues examined in this study, this was the only sample to be obtained from a site other than the bladder.

Discussion

PSCA is a cell surface antigen that has potential utility in the diagnosis and treatment of prostate cancer. In the present study, we demonstrate that PSCA is also expressed in normal bladder and in a large percentage of human TCCs. Interestingly, we found that expression in normal bladder is limited to the umbrella cell layer, a differentiated cell type believed to play a role in maintaining the integrity of the urothelium. Although the function of PSCA remains unknown, homologues of PSCA have been implicated in homotypic cell adhesion (15). Bahrenberg *et al.* (16) recently reported that PSCA expres-

Table 2 *Comparison of IHC staining intensity, density, and overall score by grade*

	Intensity					Density							Overall score			
	0	1+	2+	3+	Mean ± SD	0	1	2	3	4	5	Mean ± SD	0–3	4–7	8–15	Mean ± SD
STCC (stages T_a–T_1)																
Grade 2, n = 13 (100%)	1 (7.7)	4 (31)	8 (62)	0	1.5 ± 0.6	1 (8)	0	2 (15)	0	3 (23)	7 (54)	3.9 ± 1.6	3 (23)	2 (15)	8 (62)	6.8 ± 3.5[a]
Grade 3, n = 25 (100%)	0	2 (8)	11 (44)	12 (48)	2.4 ± 0.6	0	0	1 (4)	3 (12)	3 (3)	18 (72)	4.5 ± 0.9	2 (8)	2 (8)	21 (84)	11.2 ± 4.1[a]
ITCC (stages T_2–T_4)																
Grade 1, n = 1 (100%)	1 (100)	0	0	0	0.0 ± 0.0	1 (100)	0	0	0	0	0	0.0 ± 0.0	1 (100)	0	0	0.0 ± 0.0
Grade 2, n = 12 (100%)	8 (67)	2 (17)	1 (8)	1 (8)	0.6 ± 1.0	8 (67)	0	2 (17)	1 (8)	0	1 (8)	1.0 ± 1.7	10 (83)	1 (8.3)	1 (8.3)	2.1 ± 4.4[a]
Grade 3, n = 52 (100%)	14 (27)	8 (15)	11 (21)	19 (37)	1.6 ± 1.2	14 (27)	0	8 (15)	12 (23)	8 (15)	10 (19)	2.6 ± 1.8	17 (33)	17 (33)	18 (34)	5.9 ± 5.1[a]

[a] Two-way ANOVA using staining overall score as the dependent variable and grade and histopathologic subgroup (*i.e.*, non-ITCC and ITCC) as factors was performed. Within each subgroup, higher grade correlated with higher overall score ($P = 0.0002$).

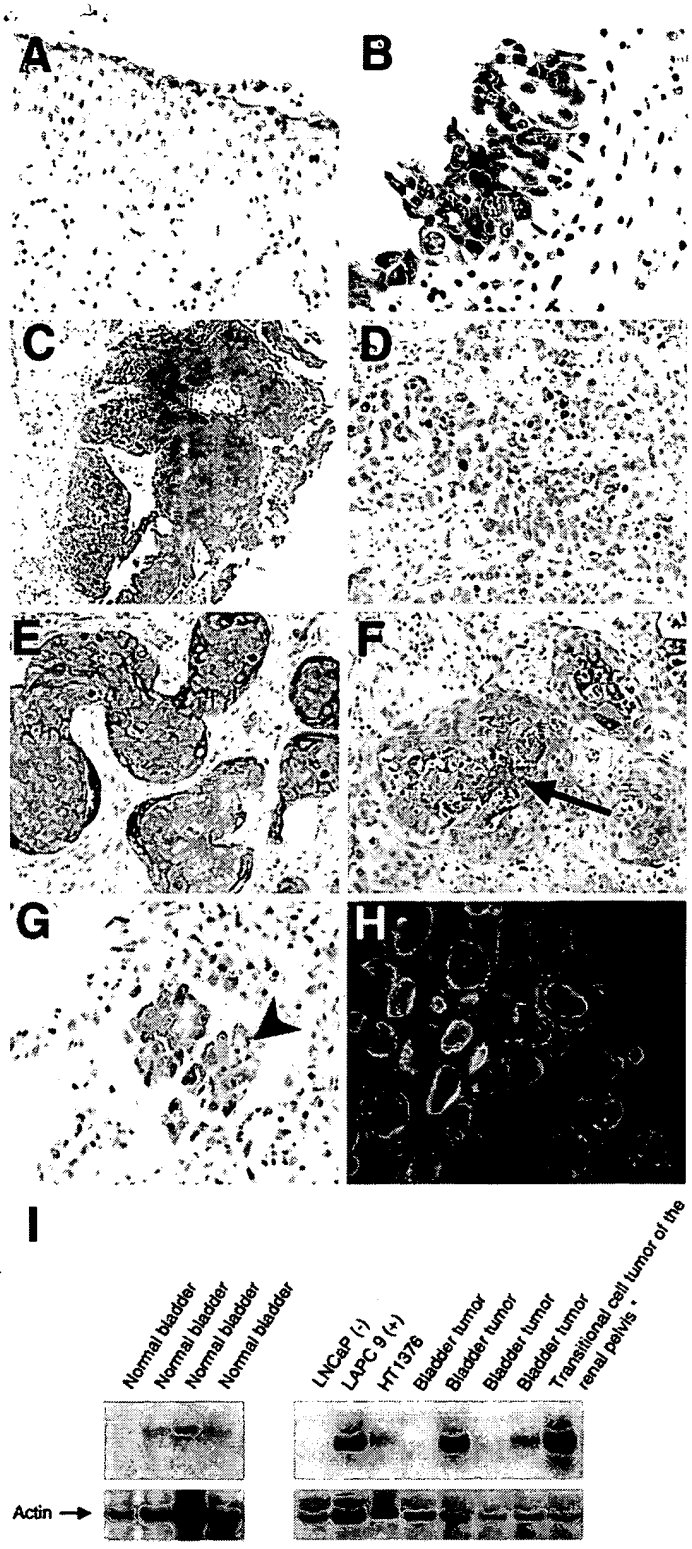

Fig. 1. IHC staining and Northern analysis of PSCA in benign and malignant urothelial tissue. All of the sections were stained with PSCA monoclonal antibody 1G8. *Brown* color, positive stain. *A*, normal bladder tissue displayed weak, superficial staining. *B*, intense, homogeneous staining was seen in CIS specimens. Positive staining can be seen in all layers of neoplastic urothelium. *C*, nonmuscle invasive superficial TCCs showed prominent PSCA expression. *D* and *E*, high-grade muscle-invasive transitional cell tumors displayed moderate-to-strong PSCA immunoreactivity. PSCA mRNA analysis from the tumor in *E* may be seen in *I*, *Lane with asterisk*. *F*, high-grade muscle-invasive tumors occasionally contained areas of squamous cell differentiation that were hot spots of PSCA expression (*arrow*). *G*, several metastatic tumors, as seen in this example from a metastatic lesion to the lung, displayed significant PSCA expression. Note the intensely positive island of metastatic transitional cells (*arrowhead*) surrounded by normal lung parenchyma. *H*, immunofluorescent staining of the cell line HT1376 shows PSCA expression (*green, fluorescein*) exclusively at the cell surface. Positive staining was seen in only a subpopulation of cells. The nuclei are counterstained with propidium iodide (*red*). *I*, Northern blot analysis of PSCA expression in normal and malignant urothelium. Three of four normal bladder specimens showed weak PSCA mRNA expression. The human bladder cancer line HT1376 displayed weak PSCA expression. Two of five muscle-invasive TCCs expressed high levels of PSCA mRNA when compared with normal bladder. Weak expression of PSCA was seen in an additional bladder tumor. All of the transitional cell tumors were muscle invasive. The human prostate cancer cell lines LNCaP and LAPC 9 were used as negative and positive controls, respectively. See "Results" for discussion of sample that was labeled with asterisk.

Table 3 *One-way ANOVA*:[a] mean overall score versus histopathology

		Absolute difference of the mean overall score *P*			
	Normal	Low-grade dysplasia	CIS	STCC	ITCC
Dysplasia	2.5				
	0.782				
CIS	12.1	9.6			
	0.001[b]	0.001[b]			
STCC	8.6	6.0	3.5		
	0.001[b]	0.001[b]	0.046[b]		
ITCC	4	1.5	8.0	4.5	
	0.001[b]	0.992	0.001[b]	0.001[b]	
Metastases	5.2	2.6	6.9	3.4	1.1
	0.066	0.968	0.005[b]	0.581	1.000

[a] One-way ANOVA-Sidak multiple comparison procedure using mean immunostaining overall score as the dependent variable and grade and histopathological subgroup as factors.

[b] Statistically significant.

sion may be regulated by cell-cell contact. It will be important to determine whether PSCA plays a role in normal bladder function and whether PSCA expression is altered in disease states such as interstitial cystitis, which are characterized by a breakdown in the normal urothelial barrier.

PSCA expression was strongest and most common in superficial, nonmuscle invasive tumors. PSCA expression was detected in all but one case of superficial transitional cell cancer and all cases of CIS. Similarly, 90% of CIS specimens and 76% of nonmuscle invasive cancers overexpressed PSCA. In contrast, 35% of invasive cancers had no detectable expression of PSCA, and only 32% overexpressed it. It is unlikely that the failure to detect expression in many invasive cancers was antibody- or antigen-related, because a similar percentage of tumors had no expression on Northern analysis. It is also unlikely that loss of expression can be attributed to a loss of differentiation in muscle-invasive tumors, as was recently proposed by Bahrenberg *et al.*, because the highest levels of PSCA expression were found in CIS, a poorly differentiated and aggressive lesion. Similarly, PSCA expression was higher in poorly differentiated nonmuscle invasive tumors than in differentiated superficial ones. Although it is possible that the loss of PSCA expression is associated with invasion, 32% of invasive and 43% of metastatic tumors overexpressed PSCA. Additional studies will be needed to understand PSCA gene regulation during the process of bladder cancer invasion.

As noted above, increasing levels of PSCA expression correlated with increasing tumor grade in both superficial and muscle-invasive tumors. One possible mechanism for PSCA overexpression is that it may result from PSCA gene amplification. PSCA maps distal to the *MYC* oncogene on chromosome 8q24.2 (8). In prostate cancer, PSCA overexpression is associated with PSCA and *MYC* coamplification (17). Christoph *et al.* (18) and Sauter *et al.* (19) have shown that low-level *MYC* amplification is a common feature of both nonmuscle-invasive and invasive bladder cancers and correlates with worsening tumor grade. Fluorescent *in situ* studies of bladder cancers should clarify whether PSCA overexpression is caused by gene amplification. Also, it will be interesting to determine whether PSCA overexpression correlates with *MYC* amplification.

We observed intense PSCA staining in poorly differentiated tran-

sitional cell tumors with squamous features, which suggests that PSCA expression may be associated with squamous differentiation. We recently created a transgenic mouse model using the PSCA promoter to drive green fluorescent protein (GFP) expression.[5] In this model, we noted GFP expression in keratinized skin as well as in the adult urothelium of a single founder line. Likewise, Bahrenberg *et al.* (16) recently reported that PSCA is expressed by keratinocytes in tissue culture, although we have not seen PSCA expression in adult skin. These observations suggest that PSCA may be a marker of both transitional and squamous differentiation patterns in the bladder and may provide another possible mechanism whereby PSCA expression levels may increase in poorly differentiated bladder tumors.

Our results differ somewhat from those reported recently by Bahrenberg *et al.* (16). Similar to Bahrenberg, we saw dramatically increased levels of PSCA expression in superficial tumors. Likewise, we saw lesser overall expression of PSCA in muscle-invasive tumors. However, unlike Bahrenberg, we found that a majority of invasive and metastatic tumors expressed PSCA and >30% overexpressed it. In addition, although Bahrenberg *et al.* concluded that PSCA expression was a marker of differentiated bladder tumors, we found that PSCA expression was highest in poorly differentiated invasive and superficial tumors. One possible reason for these differences is that the study of Bahrenberg *et al.* did not include cases of CIS or high-grade superficial cancer. Also, their study was small and included only two well-differentiated (presumably superficial) and eight poorly differentiated (presumably invasive) cancers.

Another difference between this study and that of Bahrenberg is that we primarily evaluated PSCA protein expression, whereas the latter looked at mRNA levels exclusively. Importantly, Bahrenberg *et al.* reported finding several PSCA splice variants. One of these, denoted Δ(1–17), results in the substitution of the 17-amino-acid signal sequence of exon 1 with a new 31-amino-acid sequence. This splice variant is transcribed and would be predicted to contain the epitope in exon 2 that is recognized by PSCA monoclonal antibody 1G8. It is possible, therefore, that, in at least some instances, our immunostaining would have detected a PSCA splice variant. That said, we did not detect any aberrant messages by Northern analysis, and we were clearly able to correlate PSCA mRNA and protein expression in a patient with muscle-invasive TCC. Also, our Northern analysis was consistent with the IHC data, showing expression in 60% of cases and overexpression in 40%.

Confocal microscopy studies demonstrated cell surface expression of PSCA in a subset of cells from the bladder cancer cell line HT1376. IHC analysis of PSCA shows cell surface as well as apparent cytoplasmic staining of PSCA in benign and malignant transitional epithelia (Fig. 1, *A–G*). One possible explanation for this is that anti-PSCA antibody can recognize PSCA peptide precursors that reside in the cytoplasm. In addition, it is possible that the positive staining that appears in the cytoplasm is actually from the overlying cell membrane. We have observed similar results with prostate cancer *vis-a-vis* confocal/immunofluorescent and IHC data (9). These data seem to indicate that PSCA is a novel cell surface marker for TCC.

Our results suggest that PSCA may have a number of potential uses in the diagnosis and treatment of human TCC. Because neoplastic transitional cells are naturally sloughed from the bladder, differential expression of PSCA in voided samples might form the basis for a diagnostic test. Detection of PSCA on the surface of circulating cells in blood and bone marrow might be useful in identifying some patients with bladder cancer micrometastases, particularly those whose primary tumors express PSCA. We have recently shown that immunomagnetic beads conjugated with PSCA monoclonal antibodies can detect as few as 1 in 10^7 PSCA-positive cells in bone marrow.[6] Similar approaches have been tested using uroplakin II as a marker of metastatic transitional cells (7).

PSCA might also be a valuable target for bladder cancer immunotherapy. One potential approach is to use monoclonal antibodies intravesically to treat localized disease (particularly CIS) or systemically to treat advanced disease. We have recently shown that naked PSCA antibodies inhibit prostate cancer tumorigenesis and metastasis, which suggests that they may have utility in TCC as well (10). Similar approaches have been reported in bladder cancer using antibodies against EGFR systemically and against the MUC 1 mucin antigen intravesically (5, 20). PSCA might also be useful for local or systemic vaccine approaches. Dannull *et al.* (11) recently demonstrated that amino acids 14–22 of PSCA can elicit a PSCA-specific T-cell response in a patient with prostate cancer. Given the profound immunosensitvity of bladder cancer to nonspecific agents such as BCG, this is a particularly promising approach, especially in the local setting.

In summary, we have shown in this study that PSCA is expressed by a majority of superficial and muscle-invasive transitional cell tumors. Furthermore, PSCA is overexpressed in a majority of superficial tumors and a significant percentage of invasive and metastatic transitional cell cancers. These results suggest that PSCA may be a valuable target for bladder cancer diagnosis and therapy.

References

1. Landis, S. H., Murray, T., Bolden, S., and Wingo, P. A. Cancer statistics, 1999. Cancer J. Clin., *9:* 8–31, 1999.
2. Pow-Sang, J. M., and Seigne, J. D. Contemporary management of superficial bladder cancer. Cancer Control, *7:* 335–339, 2000.
3. Messing, E. M., and Catalona, W. Urothelial tumors of the urinary tract. *In:* P. W. Walsh, A. B Retik, E. D. Vaughan, and A. J. Wein (eds.), Campbell's Urology, pp. 2327–2391. Philadelphia: W. B. Saunders Company, 1998.
4. Koenig, F., Klaus, J., Schorr, D., and Loening, S. A. Urinary markers of malignancy. Clin. Chim. Acta., *297:* 191–205, 2000.
5. Perrotte, P., Matsumoto, T., Inoue, K., Kuniyasu, H., Eve, B. Y., Hicklin, D. J., Radinsky, R., and Dinney, C. P. Anti-epidermal growth factor receptor antibody C225 inhibits angiogenesis in human transitional cell carcinoma growing orthotopically in nude mice. Clin. Cancer Res., *5:* 257–265, 1999.
6. Wu, R., L., Osman, I., Wu, X. R., Lu, M. L., Zhang, Z. F., Liang, F. X., Hamza, R., Scher, H., Cordon-Cardo, C., and Sun, T. T. Uroplakin II is expressed in transitional cell carcinoma but not in bilharzial bladder squamous cell carcinoma: alternative pathways of bladder epithelial differentiation and tumor formation. Cancer Res., *58:* 1291–1297, 1998.
7. Li, S. M., Zhang, Z. T., Chan, S., McLenan, O., Dixon, C., Taneja, S., Lepor, H., Sun, T. T., and Wu, X. R. Detection of circulating uroplakin-positive cells in patients with transitional cell carcinoma of the bladder. J. Urol., *162:* 931–935, 1999.
8. Reiter, R. E., Gu, Z., Watabe, T., Thomas, G., Szigeti, K., Davis, E., Wahl, M., Nisitani, S., Yamashiro, J., Le Beau, M. M., Loda, M., and Witte, O. Prostate stem cell antigen: a cell surface marker overexpressed in prostate cancer. Proc. Natl. Acad. Sci. USA, *95:* 1735–1740, 1998.
9. Gu, Z., Thomas, G., Yamashiro, J., Shintaku, I. P., Dorey, F., Raitano, A., Witte, O. N., Said, J. W., Loda, M., and Reiter, R. E. Prostate stem cell antigen (PSCA) expression increases with high Gleason score, advanced stage and bone metastasis in prostate cancer. Oncogene, *19:* 1288–1296, 2000.
10. Saffran, D. C., Raitano, A. B., Hubert, R. S., Witte, O. N., Reiter, R. E., and Jakobovits, A. Anti-PSCA monoclonal antibodies inhibit tumor growth and metastasis formation and prolong the survival of mice bearing human prostate cancer xenografts. Proc. Natl. Acad. Sci. USA, in press, 2001.
11. Dannull, J., Diener, P., Prikler, L., Furstenberger, G., Cerny, T., Schmid, U., Ackermann, D. K., and Groettrup, M. Prostate stem cell antigen is a promising candidate for immunotherapy of advanced prostate cancer. Cancer Res., *60:* 5522–5528, 2000.
12. Said, J. W., Pinkus, J. L., Shintaku, I. P., deVos, S., Matsumura, F., Yamashiro, S., and Pinkus, J. S. Alterations in fascin expressing germinal center dendritic cells in neoplastic follicles of B-cell lymphomas. Mod. Pathol., *11:* 1–5, 1998.
13. Hanas, J. S., Lerner, M. R., Lightfoot, S. A., Raczkowski C., Kastens, D. J., Brackett, D. J., and Postier, R. G. Expression of the cyclin-dependent kinase inhibitor p21

[5] T. Watabe, R. E. Reiter, M. Lin, A. A. Donjacour, G. R. Cunha, and O. N. Witte. *In situ* visualization of a growth responsive subset of epithelial cells associated with prostate development, regeneration and tumorigenesis, submitted for publication.

[6] Z. Gu and R. E. Reiter, unpublished data.

(WAF1/CIP1) and p53 tumor suppressor in dysplastic progression and adenocarcinoma in Barrett esophagus. Cancer (Phila.), *86:* 756–763, 1999.
14. Sambrook, J., Fritsch, E. F., and Maniatis, T. Extraction, purification, and analysis of messenger RNA from eukaryotic cells. *In:* Molecular Cloning, Ed. 2, pp. 7.3–7.87. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989.
15. Brakenhoff, R. H., Gerretsen, M., Knippels, E. M. C., van Dijk, M., van Essen, H., Weghuis, D. O., Sinke, R. J., Snow, G. B., and van Dongen, G. A. M. S. The human E48 antigen, highly homologous to the murine Ly-6 antigen ThB, is a GPI-anchored molecule apparently involved in keratinocyte cell-cell adhesion. J. Cell Biol., *129:* 1677–1689.
16. Bahrenberg, G., Brauers, A., Joost, H., and Jaske, G. Reduced expression of PSCA, a member of the Ly-6 family of cell surface antigens, in bladder, esophagus, and stomach tumors. Biochem. Biophys. Res. Comm., *275:* 783–788, 2000.
17. Reiter, R. E., Sato, I., Thomas, G., Qian, J., Gu, Z., Watabe, T., Loda, M., and Jenkins, R. B. Coamplification of prostate stem cell antigen (PSCA) and myc in locally advanced prostate cancer. Genes Chromosomes Cancer, *00:* 1–9, 1999.
18. Christoph, F., Schmidt, B., Schmitz-Drager, B. J., and Schulz, W. A. Over-expression and amplification of the c-*myc* gene in human urothelial carcinoma. Int. J. Cancer, *84:* 169–173, 1999.
19. Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F, M. C-myc copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. Am. J. Pathol., *146:* 1131–1139, 1995.
20. Hughes, O. D., Bishop, M. C., Perkins, A. C., Wastie, M. L., Denton, G., Price, M. R., Frier, M., Denley, H., Rutherford, R., and Schubiger, P. A. Targeting superficial bladder cancer by the intravesical administration of copper-67 labeled anti-MUC1 mucin monoclonal antibody C595. J. Clin. Oncol., *18:* 363–370, 2000.

Exhibit C

FUNCTION

BestFit makes an optimal alignment of the best segment of similarity between two sequences. Optimal alignments are found by inserting gaps to maximize the number of matches using the *local homology* algorithm of Smith and Waterman.

DESCRIPTION

BestFit inserts gaps to obtain the optimal alignment of the best region of similarity between two sequences, and then displays the alignment in a format similar to the output from Gap. The sequences can be of very different lengths and have only a small segment of similarity between them. You could take a short RNA sequence, for example, and run it against a whole mitochondrial genome.

SEARCHING FOR SIMILARITY

BestFit is the most powerful method in the Wisconsin Sequence Analysis Package™ for identifying the best region of similarity between two sequences whose relationship is unknown.

EXAMPLE

The sequence gamma.seq contains an Alu family sequence somewhere in the first 500 bases. alu.seq contains a generic human Alu family repeat. The two sequences are aligned and the best segment of similarity is found with BestFit.

```
% bestfit

BESTFIT of what sequence 1 ?  gamma.seq

                Begin (* 1 *) ?
              End (* 11375 *) ?  500
            Reverse (* No *) ?

to what sequence 2 (* gamma.seq *) ?  alu.seq

                Begin (* 1 *) ?
              End (*   207 *) ?
            Reverse (* No *) ?

What is the gap creation penalty (* 5.00 *) ?

What is the gap extension penalty (* 0.30 *) ?

What should I call the paired output display file (* gamma.pair *)

Aligning ..........-..

        Gaps:     3
     Quality: 129.3
Quality Ratio: 0.625
 % Similarity: 84.466
      Length:   209

%
```

OUTPUT

Here is the output file. Notice how BestFit finds and displays only the best segments of similarity:

```
 BESTFIT of: gamma.seq  check: 6474  from: 1  to: 500

Human fetal beta globins G and A gamma
from Shen, Slightom and Smithies,  Cell 26; 191-203.
Analyzed by Smithies et al. Cell 26; 345-353.

 to: alu.seq  check: 4238  from: 1  to: 207

HSREP2 from the EMBL data library
Human Alu repetitive sequence located near the insulin gene
Dhruva D.R., Shenk T., Subramanian K.N.; "Integration in vivo into
Simian virus 40 DNA of a sequence that resembles a certain family of
genomic interspersed repeated sequences"; Proc. Natl. Acad. Sci. USA
77:4514-4518(1980).  . . .

 Symbol comparison table: Gencoredisk:[Gcgcore.Data.Rundata]Swgapdna.Cmp
 CompCheck: 5234

        Gap Weight:  5.000      Average Match:  1.000
     Length Weight:  0.300   Average Mismatch: -0.900

           Quality:  129.3              Length:    209
             Ratio:  0.625               Gaps:      3
Percent Similarity: 84.466   Percent Identity: 84.466

gamma.seq x alu.seq       June 20, 1994 15:15

  137 AGACCAACCTGGCCAACATGGTGAAATCCCATCTCTAC.AAAAATACAAA 185
      |||||| |||||||||||||||||||  ||||||||||  ||||||||||
    1 AGACCAGCCTGGCCAACATGGTGAAACTCCATCTCTACTGAAAATACAAA 50

  186 AATTAGACAGGCATGATGGCAAGTGCCTGTAATCCCAGCTACTTGGGAGG 235
      ||||||  |||||||| ||    |||||||  ||||||||||||||  |||||
   51 AATTAGCCAGGCATGGTGATGCGTGCCTGGAATCCCAGCTACTTAGGAGG 100

  236 CTGAGGAAGGAGAATTGCTTGAACCTGGAAGGCAGGAGTTGCAGTGAGCC 285
      |||||  ||  |||||  |||  ||||  |  |||  |  ||||||||||||
  101 CTGAGACAGAAGAATCCCTTAAACCAAG.AGGTGGAGGTTGCAGTGAGCC 149

  286 GAGATCATACCACTGCACTCCAGCCTGGGTGACAGAACAAGACTCTGTCT 335
      ||||||  ||  ||||||||||||||  |||||||||  ||||||  |||
  150 GAGATCGCACGGCTGCACTCCAGCCT.GGTGACAGAGCGAGACTCCATCT 198

  336 CAAAAAAAA 344
      |||||||||
  199 CAAAAAAAA 207
```

RELATED PROGRAMS

When you want an alignment that covers the whole length of both sequences, use Gap. When you are trying to find only the best segment of similarity between two sequences, use BestFit. PileUp creates a multiple sequence alignment of a group of related sequences, aligning the whole length of all sequences. DotPlot displays the entire surface of comparison for a comparison of two sequences. GapShow displays the pattern of differences between two aligned sequences. PlotSimilarity plots the average similarity of two or more aligned sequences at each position in the alignment. Pretty displays alignments of several sequences. LineUp is an editor for editing multiple sequence alignments. CompTable helps generate scoring matrices for peptide comparison.

ALGORITHM

BestFit uses the *local homology* algorithm of Smith and Waterman (Advances in Applied Mathematics 2; 482-489 (1981)) to find the best segment of similarity between two sequences. BestFit reads a scoring matrix that contains values for every possible GCG symbol match (see the LOCAL DATA FILES topic below). The program uses these values to construct a path matrix that represents the entire surface of comparison with a score at every position for the best possible alignment to that point. The *quality* score for the best alignment to any point is equal to the sum of the scoring matrix values of the matches in that alignment, less the gap creation penalty times the number of gaps in that alignment, less the gap extension penalty times the total length of all gaps in that alignment. The gap creation and gap extension penalties are set by you. If the best path to any point has a negative value, a zero is put in that position.

After the path matrix is complete, the highest value on the surface of comparison represents the end of the best region of similarity between the sequences. The best path from this highest value backwards to the point where the values revert to zero is the alignment shown by BestFit. This alignment is the best segment of similarity between the two sequences.

For nucleic acids, the default scoring matrix has a *match* value of 1.0 for each identical symbol comparison and -0.90 for each non-identical comparison (not considering nucleotide ambiguity symbols for this example). The *quality* score for a nucleic acid alignment can, therefore, be determined using the following equation:

```
Quality = 1.0 x TotalMatches +  -0.90 x TotalMismatches
                  - (GapCreationPenalty x GapNumber)
          - (GapExtensionPenalty x TotalLengthOfGaps)
```

The *quality* score for a protein alignment is calculated in a similar manner. However, while the default nucleic acid scoring matrix has a single value for all non-identical comparisons, the default protein scoring matrix has different values for the various non-identical amino acid comparisons. The *quality* score for a protein alignment can therefore be determined using the following equation (where $Total_{AA}$ is the total number of A-A (Ala-Ala) matches in the alignment, $CmpVal_{AA}$ is the value for an A-A comparison in the scoring matrix, $Total_{AB}$ is the total number of A-B (Ala-Asx) matches in the alignment, $CmpVal_{AB}$ is the value for an A-B comparison in the scoring matrix, ...):

$$
\begin{aligned}
Quality = {} & CmpVal_{AA} \times Total_{AA} \\
& + CmpVal_{AB} \times Total_{AB} \\
& - CmpVal_{AC} \times Total_{AC} \\
& \\
& \\
& - CmpVal_{zz} \times Total_{zz} \\
& - (GapCreationPenalty \times GapNumber) \\
& - (GapExtensionPenalty \times TotalLengthOfGaps)
\end{aligned}
$$

For a more complete discussion of scoring matrices, see the **Data Files** manual.

CONSIDERATIONS

BestFit Always Finds Something

BestFit always finds an alignment for any two sequences you compare -- even if there is no significant similarity between them! You must evaluate the results critically to decide if the segment shown is not just a random region of relative similarity.

The Segments Shown Obscure Alternative Segments

BestFit only shows one segment of similarity, so if there are several, all but one is obscured. You can approach this problem with graphic matrix analysis (see the Compare and DotPlot programs). Alternatively, you can run BestFit on ranges outside the ranges of similarity found in earlier runs to bring other segments out of the shadow of the best segment.

The Best Fit is Only One Member of a Family

Like all fast gapping algorithms, the alignment displayed is a member of the family of best alignments. This family may have other members of equal quality, but will not have any member with a higher quality. The family is usually significantly different for different choices of gap creation and gap extension penalties. See the CONSIDERATIONS topic in the entry for the Gap program in the **Program Manual** to learn more about how to assign gap creation and gap extension penalties.

The Surface of Comparison

The magnitude of the computer's job is proportional to the area of the surface of comparison. That area is determined by the product of the lengths of the two sequences compared. BestFit can evaluate a surface of up to 3.5 million elements. This surface would be large enough to compare two sequences approximately 1,870-symbols long, or one sequence 200-symbols long with another sequence 17,500-symbols long. When you have much longer sequences that are known to align well, you can use the command-line option `-LIMit` to use the surface more efficiently.

The Public Scoring Matrix for Nucleic Acid Comparisons is Very Stringent

The scoring matrix swgapdna.cmp penalizes mismatches -0.9 so the segments found may be very brief. This penalty means that the alignment cannot be extended by three bases to pick one extra match. The scoring matrix used by Smith and Waterman, when local alignments were first described, used -0.333 for the mismatch penalty. You can use Fetch to copy randomdna.cmp and rename it swgapdna.cmp to use these values, or use nwsgapdna.cmp, which has no mismatch penalty at all.

Rapid Alignment

When possible, BestFit tries to find the optimal alignment very quickly. If this rapid alignment is not unambiguously optimal, BestFit automatically realigns the sequences to calculate the optimal alignment. When this occurs, the monitor of alignment progress on your terminal screen (`Aligning...`) is displayed twice for a single alignment.

ALIGNING LONG SEQUENCES

This program can align very long sequences if you know roughly where the alignment of interest begins. Run the program with the command line option `-LIMit`. Then set the starting coordinates for each sequence near the point where the alignment of interest begins and set gap shift limits on each sequence. The program then aligns the sequences from your starting point such that the sequences do not get out of phase by more than the gap shift limits you have set. If you started both sequences at

base number one and set the gap shift limit for sequence one to 100 and for sequence two to 50, then base 350 in sequence one could not be gapped to any base outside of the range from 300 to 450 on sequence two.

If you omit `-LIMit` on the command line, the program automatically sets gap shift limits if they are needed to allow the alignment of long sequences to proceed. In this case, the program limits the total length of gaps that can be inserted into each sequence and calculates the best alignment within this incomplete, or *limited*, surface of comparison. The program then performs a calculation to determine whether the alignment could possibly be improved if there were no restriction on the total length of gaps in each sequence. If the program cannot rule out this possibility, it displays the message `*** Alignment is not guaranteed to be optimal ***`. Because the criteria used in the calculation for guaranteeing an optimal alignment are very stringent, a limited alignment often may be optimal even if this message is displayed. In any event, the program continues to completion.

EVALUATING ALIGNMENT SIGNIFICANCE

This program can help you evaluate the significance of the alignment, using a simple statistical method, with the `-RANdomizations` command line option. The second sequence is repeatedly shuffled, maintaining its length and composition, and then realigned to the first sequence. The average alignment score, plus or minus the standard deviation, of all randomized alignments is reported in the output file. You can compare this average *quality* score to the quality score of the actual alignment to help evaluate the significance of the alignment. The number of randomizations can be specified along with the `-RANdomizations` command line qualifier; the default is 10.

The score of each randomized alignment is reported to the screen. You can use `<Ctrl>C` to interrupt the randomizations and output the results from those randomized alignments that have been completed.

By ignoring the statistical properties of biological sequences, this simple Monte Carlo statistical method may give misleading results. Please see Lipman, D.J, Wilbur, W.J., Smith, T.F., and Waterman, M.S. (Nucl. Acids Res. 12; 215-226 (1984)) for a discussion of the statistical significance of nucleic acid similarities.

ALIGNMENT METRICS

BestFit and Gap display four figures of merit for alignments: Quality, Ratio, Identity, and Similarity.

The Quality (described above) is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the *similarity threshold*. This threshold is also used by the display procedure to decide when to put a ':' (colon) between two aligned symbols. You can reset it from the command line with the second optional parameter of `-PAIr`. For instance, the expression `-PAIr=1.0,0.5` would set the similarity threshold to 0.5.

The similarity and identity metrics are not optimized by alignment programs so they should not be used to compare alignments.

PEPTIDE SEQUENCES

If your input sequences are peptide sequences, this program uses a scoring matrix with matches scored as 1.5 and mismatches scored according to the evolutionary distance between the amino acids as measured by Dayhoff and normalized by Gribskov (Gribskov and Burgess Nucl. Acids Res. 14(16); 6745-6763 (1986)).

RESTRICTIONS

Input sequences may not be more than 30,000-symbols long. This program cannot evaluate a surface of comparison larger than 5.5 million elements. A 200 x 27,500 comparison is possible, as well as a 2,300 x 2,300 comparison. See the ALIGNING LONG SEQUENCES topic for help in aligning long sequences that would normally exceed the maximum surface of comparison. You can also ask your system manager to increase the maximum surface of comparison if your system has enough virtual memory.

SEQUENCE TYPE

The function of BestFit depends on whether your input sequence(s) are protein or nucleotide. Normally the type of a sequence is determined by the presence of either `Type: N` or `Type: P` on the last line of the text heading just above the sequence itself. If your sequence(s) are not the correct type, turn to Appendix VI for information on how to change or set the type of a sequence.

COMMAND-LINE SUMMARY

All parameters for this program may be put on the command line. Use the option `-CHEck` to see the summary below and to have a chance to add things to the command line before the program executes. In the summary below, the capitalized letters in the qualifier names are the letters that you *must* type in order to use the parameter. Square brackets ([and]) enclose qualifiers or parameter values that are optional. For more information, see "Using Program Parameters" in Chapter 3, Basic Concepts: Using Programs in the **User's Guide**.

```
Minimal Syntax: % bestfit [-INfile1=]gamma.seq  [-INfile2=]alu.seq -Default

Prompted Parameters:

-BEGin1=1 -BEGin2=1       beginning of each sequence
-END1=500 -END2=207       end of each sequence
-NOREV1     -NOREV2       strand of each sequence
-GAPweight=5.0            gap creation penalty     (3.0 is protein default)
-LENgthweight=0.3         gap extension penalty    (0.1 is protein default)
[-OUTfile1=]gamma.pair    output file for alignment

Local Data Files: -DATa=swgapdna.cmp  scoring matrix for nucleic acids
                  -DATa=swgappep.cmp  scoring matrix for peptides

Optional Parameters:

-OUTfile2=gamma.gap       new sequence file for sequence 1 with gaps added
-OUTfile3=alu.gap          "     "       "   "      "    2   "   "     "
-LIMit1=499 -LIMit2=206 limit the surface of comparison
-RANdomizations[=10]      determine average score from 10 randomized
                             alignments
-PAIr=1.0,0.5,0.1         thresholds for displaying '|', ':', and '.'
-WIDth=50                 the number of sequence symbols per line
-PAGe=60                  adds a line with a form feed every 60 lines
-NOBIGGaps                suppresses abbreviation of large gaps with '.'s
-HIGhroad                 makes the top alignment for your parameters
-LOWroad                  makes the bottom alignment for your parameters
-NOSUMmary                suppresses the screen summary
```

ACKNOWLEDGEMENTS

Gap and BestFit were originally written for Version 1.0 by Paul Haeberli from a careful reading of the Needleman and Wunsch (J. Mol. Biol. 48; 443-453 (1970)) and the Smith and Waterman (Adv. Appl. Math. 2; 482-489 (1981)) papers.

Limited alignments were designed by Paul Haeberli and added to the Package for Version 3.0. They were united into a single program by Philip Delaquess for Version 4.0. Default gap penalties for protein alignments were modified according to the suggestions of Rechid, Vingron and Argos (CABIOS 5; 107-113 (1989)).

LOCAL DATA FILES

The files described below supply auxiliary data to this program. The program automatically reads them from a public data directory unless you either 1) have a data file with exactly the same name in your current working directory; or 2) name a file on the command line with an expression like `-DATa1=myfile.dat`. For more information see Chapter 4, Using Data Files in the **User's Guide**.

If the first sequence you name is a nucleic acid, BestFit uses the scoring matrix in the public file swgapdna.cmp. (SW stands for Smith and Waterman.) If the first sequence you name is a peptide sequence, BestFit reads swgappep.cmp instead. The presence of these files in your current working directory causes BestFit to read your version instead. (See the **Data Files** manual for more information about scoring matrices.)

OPTIONAL PARAMETERS

The parameters and switches listed below can be set from the command line. For more information, see "Using Program Parameters" in Chapter 3, Basic Concepts: Using Programs in the **User's Guide**.

`-LIMit1=20 and -LIMit2=20`

let you set *gap shift limits* for each sequence. When you already know of a long similarity between two sequences you can "zip" them together using this mode. The beginning coordinates for each sequence must be near the beginning of the alignment you want to see. The alignment continues so that gaps inserted do not require the sequences to get out of step by more than the gap shift limits. You can align very long sequences rapidly. The surface of comparison is still limited to 3.5 million. The size of a comparison can be predicted by multiplying the average length of the two sequences by the sum of the two shift limits.

If you add `-LIMit` to the command line without any qualifier value, the program prompts you to enter gap shift limits for each sequence.

`-RANdomizations=10`

reports the average alignment score and standard deviation from 10 randomized alignments in which the second sequence is repeatedly shuffled, maintaining the length and composition of the original sequence, and then aligned to the first sequence. You can use the optional parameter to set the number of randomized alignment to some number other than 10.

`-OUTfile2=seqname1.gap -OUTfile3=seqname2.gap`

This program can write three different output files. The first displays the alignment of sequence one with sequence two. The second is a new sequence file for sequence one, possibly expanded by gaps to make it align with sequence two. The third, like the second, is a new sequence file for sequence two, possibly expanded by gaps to make it align with sequence one. The program writes only the first file unless there are output file options on the command line. If there are any output files named on the command line, *only* those output files are written. If you add

-OUT to the command line without any qualifying filename, then the program will write the second and third output files after prompting you for their names.

Aligned sequences (in sequence files) can be displayed with GapShow. Their similarity can be displayed with PlotSimilarity.

-PAIr=1.0,0.5,0.1

The paired output file from this program displays sequence similarity by printing one of three characters between similar sequence symbols: a pipe character(|), a colon (:), or a period (.). Normally a pipe character is put between symbols that are the same, a colon is put between symbols whose comparison value is greater than or equal to 0.50, and a period is put between symbols whose comparison value is greater than or equal to 0.10. You can change these *match display thresholds* from the command line. The three parameters for -PAIr are the display thresholds for the pipe character, colon, and period. The match display criterion for a pipe character changes from symbolic identity (the default) to the quantitative threshold you have set in the first parameter. A pipe character will no longer be inserted between identical symbols unless their comparison values are greater than or equal to this threshold. If you still want a pipe character to connect identical symbols, use x instead of a number as the first parameter. (See the **Data Files** manual for more information about scoring matrices.)

-PAGe=64

When you print the output from this program, it may cross from one page to another in a frustrating way -- especially when you print on individual sheets. This option adds form feeds to the output file in order to try to keep clusters of related information together. You can set the number of lines per page by supplying a number after the -PAGe qualifier.

-WIDth=50

puts 50 sequence symbols on each line of the output file. You can set the width to anything from 10 to 150 symbols.

-NOBIGGaps

suppresses large gap abbreviations, showing all the sequence characters across from large gaps. Usually, gaps that extend one sequence by more than one complete line of output are abbreviated with three dots arranged in a vertical line.

-LOWroad and -HIGhroad

The insertion of gaps is, in many cases, arbitrary, and equally optimal alignments can be generated by inserting gaps differently. When equally optimal alignments are possible, this program can insert the gaps differently if you select either the -LOWroad or the -HIGhroad options. Here are examples for the alignment of GACCAT with GACAT with different parameters.

```
For:        Match = 1.0        MisMatch = -0.9
      Gap weight = 1.0   Length Weight =  0.0

LowRoad:   1 GACCAT 6
             :  |||      Quality = 4.0
           1 GA.CAT 5

HighRoad:  1 GACCAT 6
             ||| ||      Quality = 4.0
           1 GAC.AT 5
```

```
For:        Match = 1.0        MisMatch =  0.0
       Gap weight = 3.0   Length Weight =  0.0

HighRoad:  1 GACCAT 6
             |||          Quality = 3.0
           1 GACAT. 5

LowRoad:   1 GACCAT 6
              |||         Quality = 3.0
           1 .GACAT 5
```

Essentially the *low road* shifts all of the arbitrary gaps in sequence two to the left and all of the arbitrary gaps in sequence one to the right. The *high road* does exactly the opposite. When neither *high road* nor *low road* is selected, the program tries not to insert a gap whenever that is possible and uses the high road alternative for all collisions.

-SUMmary

writes a summary of the program's work to the screen when you've used the `-Default` qualifier to suppress all program interaction. A summary typically displays at the end of a program run interactively. You can suppress the summary for a program run interactively with **-NOSUMmary**.

Use this qualifier also to include a summary of the program's work in the log file for a program run in batch.